

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
22 July 2004 (22.07.2004)

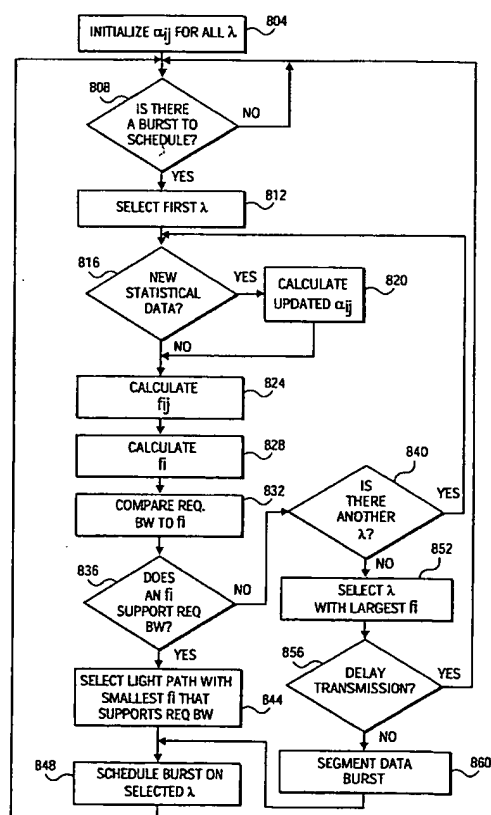
PCT

(10) International Publication Number  
WO 2004/062313 A1

- (51) International Patent Classification<sup>7</sup>: H04Q 11/00
- (21) International Application Number:  
PCT/US2003/038613
- (22) International Filing Date: 5 December 2003 (05.12.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
10/328,571 24 December 2002 (24.12.2002) US
- (71) Applicant: INTEL CORPORATION [US/US]; 2200 Mission College Boulevard, Santa Clara, CA 95052 (US).
- (72) Inventors: OVADIA, Shlomo; 1576 Willow Oaks Drive, San Jose, CA 95125 (US). MACIOCCO, Christian; 16322 SW Stahl Drive, Tigard, OR 97223 (US).
- (74) Agents: MALLIE, Michael, J. et al.; Blakely Sokoloff Taylor & Zafman, 12400 Wilshire Boulevard, 7th Floor, Los Angeles, CA 90025 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR DATA AND CONTROL PACKET SCHEDULING IN WDM PHOTONIC BURST-SWITCHED NETWORKS



(57) Abstract: A method and apparatus for scheduling data burst transmission in a wavelength-division-multiplexing photonic burst-switched network is disclosed. In one embodiment, a scheduler determines a set of weighting factors at each wavelength for each of the lightpath segments among all the network nodes. These weighting factors may be used to calculate a set of coarse-grained bandwidth allocation factors, one for each potential lightpath formed from the available lightpath segments between an ingress node and an egress node. The scheduler may schedule the transmission of the data burst on one of the lightpaths that has a coarse-grained bandwidth allocation factor which indicates that it can support the requested bandwidth. The scheduler may periodically update the set of weighting factors at each wavelength for each of the lightpath segments.

WO 2004/062313 A1



**Published:**

— with international search report

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## METHOD AND APPARATUS FOR DATA AND CONTROL PACKET SCHEDULING IN WDM PHOTONIC BURST-SWITCHED NETWORKS

**FIELD**

- 5 [0001] The present disclosure relates generally to optical networking systems, and more specifically to optical networking systems operating in photonic burst-switched modes.

**BACKGROUND**

- 10 [0002] Transmission bandwidth demands in telecommunication networks (*e.g.*, the Internet) appear to be ever increasing and solutions are being sought to support this bandwidth demand. One solution to this problem is to use fiber-optic networks, where wavelength-division-multiplexing (WDM) technology is used to support the ever-growing demand in optical networks for higher data rates.
- 15 [0003] Conventional optical switched networks typically use wavelength routing techniques, which require that optical-electrical-optical (OEO) conversion of optical signals is done at the optical switches. Performing OEO conversion at each switching node in the optical network is not only a relatively slow operation (typically about ten milliseconds), but is also
- 20 very costly, consumes a lot of power, and potentially creates a traffic bottleneck for the optical switched network. In addition, the current optical switch technologies cannot efficiently support "bursty" traffic that is often experienced in packet communication applications (*e.g.*, the Internet).
- 25 [0004] It may be possible to create a network where the OEO conversion at the switching nodes need not be performed. In such a case, the data traffic maybe electronically buffered at the ingress/egress nodes, and their transmission to egress/ingress nodes is placed upon network schedulers. Prior art bandwidth reservation mechanisms may not be
- 30 adequate for enterprise networks to support the requested bandwidths, as they did not take into account the specific network topology and available resources such as the number of active wavelengths permitted, the lightpath utilization, and many other network parameters.

**BRIEF DESCRIPTION OF THE DRAWINGS**

[0005] The present invention is illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings and in which like reference numerals refer to similar elements and in which:

5

[0006] **Figure 1** is a simplified block diagram illustrating a photonic burst switching network, according to one embodiment of the present disclosure.

10 [0007] **Figure 2** is a simplified flow diagram illustrating the operation of a photonic burst switching network, according to one embodiment of the present disclosure.

[0008] **Figure 3** is a block diagram illustrating a core switching node module for use in a photonic burst switching network, according to one embodiment of the present disclosure.

15 [0009] **Figures 4A and 4B** are diagrams illustrating the format of an optical data burst and an optical control burst for use in a photonic burst-switched network, according to one embodiment of the present disclosure.

20 [0010] **Figure 5** is a simplified block diagram architecture of certain portions of an edge router, that may be used as an ingress node or egress node, showing various features such as a traffic shaper and multiple queues, according to one embodiment of the present disclosure.

25 [0011] **Figures 6A and 6B** are diagrams illustrating a method of selecting a lightpath in a five switching node photonic burst switching network, according to one embodiment of the present disclosure.

[0012] **Figure 7** is a simplified block diagram of software modules for use in an ingress node or egress node such as the edge router shown in Figure 5, according to one embodiment of the present disclosure.

30 [0013] **Figure 8** is a flowchart diagram of a method discussed in connection with Figures 6A and 6B, according to one embodiment of the present disclosure.

**DETAILED DESCRIPTION**

[0014] The following disclosure describes techniques for reserving bandwidth and thereby assisting the scheduling of data burst transmissions in a WDM-based photonic burst-switched network. In the following description, numerous specific details such as logic implementations, software module allocation, bus signaling techniques, and details of operation are set forth in order to provide a more thorough understanding of the present invention. It will be appreciated, however, by one skilled in the art that the invention may be practiced without such specific details. In other instances, control structures, gate level circuits and full software instruction sequences have not been shown in detail in order not to obscure the invention. Those of ordinary skill in the art, with the included descriptions, will be able to implement appropriate functionality without undue experimentation. The invention is disclosed in the form of hardware and software within a photonic burst-switched network. However, the invention may be practiced in other forms of networks.

[0015] Figure 1 illustrates a photonic burst-switched network 10, according to one embodiment of the present disclosure. The term photonic burst is used herein to refer to statistically-multiplexed packets (*e.g.*, internet protocol (IP) packets, Ethernet frames) having similar routing requirements. A photonic burst typically includes a photonic control burst containing the header and other routing, scheduling, or network management information of the IP packets or any other frame and a payload containing the data segments of the packets.

[0016] This embodiment of photonic burst-switched network 10 includes an optical metropolitan area network MAN 11, local area networks (LANs) 13<sub>1</sub>-13<sub>N</sub> and a backbone optical wide area network WAN (not shown). In addition, this embodiment of optical MAN 11 includes ingress nodes 15<sub>1</sub>-15<sub>M</sub>, switching nodes 17<sub>1</sub>-17<sub>L</sub>, and egress nodes 18<sub>1</sub>-18<sub>N</sub>. Optical MAN 11 may include other ingress nodes, egress nodes, and switching nodes (not shown) that may be interconnected with the switching nodes shown in Figure 1. In this embodiment, the ingress,

egress, and switching nodes may be implemented with intelligent modules.

**[0017]** In this embodiment, ingress nodes 15<sub>1</sub>-15<sub>M</sub> may be implemented as label edge routers (LER) with the appropriate optical interface units or modules, where each is configured to receive IP packets or frames (e.g. Ethernet, SONET) from a LAN (via a label edge router (LER) in some embodiments or directly on its interfaces), and receive optical signals from switching nodes 17<sub>1</sub>-17<sub>M</sub> of optical MAN 11. In other embodiments, ingress nodes 15<sub>1</sub>-15<sub>M</sub> may be implemented as a server, or as a server farm. In addition, ingress nodes 15<sub>1</sub>-15<sub>M</sub> may be configured to transmit optical signals to switching nodes 17<sub>1</sub>-17<sub>M</sub> of optical MAN 11. In one embodiment, the ingress nodes may perform optical-electrical (OE) conversion of received optical signals, and include electronic memory to buffer the received signals until they are sent to the appropriate LAN. In another embodiment, the ingress switching nodes may perform electrical-optical (EO) conversion of the received electrical signals before they are transmitted to switching nodes 17<sub>1</sub>-17<sub>M</sub> of optical MAN 11. Embodiments of the ingress nodes are described further below.

**[0018]** Egress nodes 18<sub>1</sub>-18<sub>N</sub> may be implemented with modules that may be configured to receive optical signals from other nodes of optical MAN 11 and route them to the optical WAN or to a LAN. Egress switching node 18<sub>1</sub>-18<sub>N</sub> can also receive optical signals from the optical WAN and send them to the appropriate node of optical MAN 11. In one embodiment, egress node 18<sub>1</sub>-18<sub>N</sub> performs optical-electrical-optical (OEO) conversion of received optical signals, and includes electronic memory to buffer received signals until they are sent to the appropriate node of optical MAN 11 (or to the optical WAN). Embodiments of egress node 18<sub>1</sub>-18<sub>N</sub> are described further below.

**[0019]** Switching nodes 17<sub>1</sub>-17<sub>L</sub> may be implemented with optical switching units or modules that may be each configured to receive optical signals from other switching nodes and appropriately route the received optical signals to other switching nodes of optical MAN 11. As is

described below, the core switching nodes perform OEO conversion of optical control bursts and network management control label signals. In some embodiments, these optical control bursts and control labels are propagated only on preselected wavelengths. The preselected wavelengths  
5 do not propagate optical "data" burst (as opposed to control bursts and control labels) signals in such embodiments, even though the burst and network management control labels may include necessary information for a particular group of optical data burst signals. In another embodiment, optical control bursts, network management control labels,  
10 and optical data burst signals may be propagated on the same wavelength(s) using different encoding schemes such as different modulation formats, etc. In either approach, the optical control bursts and control labels are sent asynchronously relative to its corresponding optical data burst signals. In still another embodiment, the optical  
15 control bursts and control signals are propagated at different transmission rates as the optical data signals.

**[0020]** Although switching nodes 17<sub>1</sub>-17<sub>L</sub> may perform OEO conversion of the optical control burst signals, in this embodiment, the switching nodes do not perform OEO conversion of the optical data burst signals.  
20 Rather, core switching nodes 17<sub>1</sub>-17<sub>L</sub> perform purely optical switching of the optical data burst signals. Thus, the switching nodes can include electronic circuitry to store and process the incoming optical control bursts and network management control labels that were converted to an electronic form and use this information to configure the photonic burst  
25 switch (PBS) settings, and to properly route the optical data burst signals corresponding to the optical control bursts. The new control bursts, which replace the previous control bursts based on the new routing information, may be converted to an optical control signal, and may then be transmitted to the switching or egress nodes. Embodiments of the core  
30 switching nodes are described further below.

**[0021]** The elements of one photonic burst switching network 10 may be interconnected as follows. LANs 13<sub>1</sub>-13<sub>N</sub> may be connected to corresponding ones of ingress switching nodes 15<sub>1</sub>-15<sub>M</sub> of optical MAN 11.

[0022] Within optical MAN 11, ingress nodes 15<sub>1</sub>-15<sub>M</sub> and egress node 18 may be connected to some of switching nodes 17<sub>1</sub>-17<sub>L</sub> via optical fibers. Switching nodes 17<sub>1</sub>-17<sub>L</sub> may also be interconnected to each other via optical fibers in mesh architecture to form a relatively large number of optical paths or links between the ingress nodes, and between ingress nodes 15<sub>1</sub>-15<sub>L</sub> and egress nodes 18<sub>1</sub>-18<sub>N</sub>. Ideally, core switching nodes 17<sub>1</sub>-17<sub>L</sub> provide more than one optical path between each end point of optical MAN 11 (i.e., the ingress nodes 15<sub>1</sub>-15<sub>L</sub> and egress nodes 18<sub>1</sub>-18<sub>N</sub> are endpoints within optical MAN 11). Multiple optical paths between core switching nodes, ingress nodes, and egress nodes may enable fast reroute and protection switching when, for example, one or more nodes fail.

[0023] As described below in conjunction with Figure 2, the ingress, egress and switching nodes of optical MAN 11 may be configured to send and/or receive optical control bursts, optical data bursts, and other control signals that are wavelength multiplexed so as to propagate the optical control bursts and control labels on pre-selected wavelength(s) and optical data bursts or payloads on different preselected wavelength(s). In addition, optical data bursts may be time division multiplexed (TDM) on a given wavelength. Still further, the endpoints of optical MAN 11 may send optical control burst signals when sending data out of optical MAN 11.

[0024] Figure 2 illustrates the operational flow of photonic burst switching network 10, according to one embodiment of the present disclosure. Referring to Figures 1 and 2, photonic burst switching network 10 operates as follows.

[0025] Optical MAN 11 receives packets from LANs 13<sub>1</sub>-13<sub>N</sub>. In one embodiment, optical MAN 11 receives IP packets at ingress nodes 15<sub>1</sub>-15<sub>M</sub>. The received packets may be in electronic form rather than in optical form, or received in optical form and then converted to electronic form. In this embodiment, the ingress nodes store the received packets electronically. A block 20 represents this operation.

[0026] For clarity, the rest of the description of the operational flow of photonic burst switching network 10 focuses on the transport of information from ingress node 15<sub>1</sub> to egress node 18<sub>1</sub>. The transport of information from ingress nodes 15<sub>2</sub>-15<sub>M</sub> to one of the egress nodes 18<sub>1</sub>-18<sub>N</sub> may be substantially similar.

[0027] An optical control burst and payload (*i.e.*, an optical data burst) is formed from the received packets. In one embodiment, ingress node 15<sub>1</sub> uses statistical multiplexing techniques to form the optical data burst from the received IP (Internet Protocol) packets or frames (Ethernet, SONET, etc) stored in ingress node 15<sub>1</sub>. For example, packets received by ingress node 15<sub>1</sub> and having to pass through egress node 18<sub>1</sub> on their paths to a destination may be assembled into an optical data burst. A block 21 represents this operation.

[0028] Bandwidth may be reserved on a specific wavelength to transport the optical data burst through photonic burst switching network 10. In one embodiment, ingress node 15<sub>1</sub> reserves a time slot (*e.g.*, a TDM channel of a TDM system) in an optical data signal path through photonic burst switching network 10. Further, in one embodiment, the bandwidth is reserved for a time period sufficient to transport the optical burst from the ingress node to the egress node. For example, in some embodiments, the ingress, core, and egress nodes maintain an updated list of all used and available time slots. The time slots may be allocated and distributed over multiple wavelengths and optical fibers. Thus, a reserved time slot (also referred to herein as a TDM channel, which in differing embodiments may be either fixed-duration or variable duration TDM channels) may be in one wavelength of one fiber, or may be spread across multiple wavelengths and multiple optical fibers. A block 22 represents this operation.

[0029] When an ingress and/or egress node reserves bandwidth on a specific wavelength or when bandwidth is released after an optical data burst is transported, a network controller (not shown) updates the list. In one embodiment, the network controller and the ingress or egress nodes

perform this updating process using various burst or packet scheduling algorithms based on the available network resources and traffic patterns. The available variable-duration TDM channels, which may be periodically broadcasted to all the ingress, switching, and egress nodes, may be  
5 transmitted on the same wavelength as the optical control bursts or on a different common preselected wavelength throughout the optical network. The network controller function may reside in one of the ingress and egress nodes or may be distributed across two or more ingress and egress nodes. In this embodiment, the network controller also resides in the  
10 switching node (*e.g.*, processors 82 or 83 of Figure 8).

**[0030]** The optical control bursts, network management control labels, and optical data bursts may be then transported through photonic burst switching network 10 in the reserved variable-duration TDM channels. That is, each of the reserved time slots may have different time durations,  
15 depending on the data burst lengths. In one embodiment, ingress node 15<sub>1</sub> transmits the control burst to the next switching node along the optical label-switched path (OLSP) determined by the network controller. In this embodiment, the network controller uses a constraint-based routing protocol (*e.g.*, multi-protocol label switching MPLS) over one or  
20 more wavelengths to determine the best available OLSP to the egress node. In one embodiment, ingress node 15<sub>1</sub> then asynchronously transmits the optical data burst to the destination egress node through switching nodes 17<sub>1</sub>-17<sub>L</sub>, where little or no time delay due to buffering or OEO conversion occurs at each switching node.

25 **[0031]** In some embodiments, the switching node may perform OEO conversion of the optical control burst so that the node may extract and process the routing information contained in the control burst. Further, in some embodiments, the variable-duration TDM channel may be propagated in the same wavelengths that are used for propagating control  
30 bursts. Alternatively, the control bursts and payloads may be modulated on the same wavelength in the same optical fiber using different modulation formats. For example, optical control bursts may be transmitted using non-return-to-zero (NRZ) modulation format, while

optical payloads may be transmitted using return-to-zero (RZ) modulation format. The optical burst is transmitted from one switching node to another switching node in a similar optical MAN until the optical burst is terminated at egress node 18<sub>1</sub>. A block 23 represents this operation.

5 [0032] The operational flow at this point depends on whether the target network is an optical WAN or a LAN. A block 24 represents this branch in the operational flow.

[0033] If the target network is an optical WAN, new optical control burst and payload signals may be formed. In this embodiment, egress node 18<sub>1</sub>  
10 prepares the new optical control burst and payload signals. A block 25 represents this operation.

[0034] The new optical control burst and payload may then be transmitted to the target network (*i.e.*, WAN in this case). In this embodiment, egress node 18<sub>1</sub> includes an optical interface to transmit the  
15 optical control burst and payload to the optical WAN. A block 26 represents this operation.

[0035] However, if in block 24 the target network is a LAN, the optical data burst is disassembled to extract the IP packets or frames. In this embodiment, egress node 18<sub>1</sub> converts the optical data burst to electronic  
20 signals that egress node 18<sub>1</sub> may process to recover the data segment of each of the packets. A block 27 represents this operation.

[0036] The extracted IP data packets or frames may be processed, combined with the corresponding IP labels, and then routed to the target network (*i.e.*, LAN in this case). In this embodiment, egress node 18<sub>1</sub>  
25 forms these new IP packets. A block 28 represents this operation. The new IP packets may then be transmitted to the target network (*i.e.*, LAN) as shown in block 26.

[0037] Photonic burst switching network 10 may achieve increased bandwidth efficiency through the additional flexibility afforded by the  
30 TDM channels. Although this embodiment described above includes an optical MAN having ingress, switching, and egress nodes to couple

multiple LANs to an optical WAN backbone, in other embodiments the networks do not have to be LANs, optical MANs or WAN backbones. That is, optical MAN 11 need not service a "metropolitan area". Rather, photonic burst switching network 10 may include a number of relatively small networks that are coupled to a relatively larger network that in turn is coupled to a backbone network.

[0038] Figure 3 illustrates a module 17 for use as a switching node in photonic burst switching network 10 (Figure 1), according to one embodiment of the present disclosure. In this embodiment, module 17 includes a set of optical wavelength division demultiplexers  $30_1-30_A$ , where  $A$  represents the number of input optical fibers used for propagating payloads, control bursts, and other network resources to the module. For example, in this embodiment, each input fiber could carry a set of  $C$  wavelengths (i.e., WDM wavelengths), although in other embodiments the input optical fibers may carry differing numbers of wavelengths. Module 17 would also include a set of  $N \times N$  photonic burst switches  $32_1-32_B$ , where  $N$  is the number of input/output ports of each photonic burst switch. Thus, in this embodiment, the maximum number of wavelengths at each photonic burst switch is  $A \cdot C$ , where  $N \geq A \cdot C + 1$ . For embodiments in which  $N$  is greater than  $A \cdot C$ , the extra input/output ports may be used to loop back an optical signal for buffering.

[0039] Further, although photonic burst switches  $32_1-32_B$  are shown as separate units, they may be implemented as  $N \times N$  photonic burst switches using any suitable switch architecture. Module 17 also includes a set of optical wavelength division multiplexers  $34_1-34_A$ , a set of optical-to-electrical signal converters 36 (e.g., photo-detectors), a control interface unit 37, and a set of electrical-to-optical signal converters 38 (e.g., lasers).

[0040] The elements of this embodiment of module 17 may be interconnected as follows. Optical demultiplexers  $30_1-30_A$  may be connected to a set of  $A$  input optical fibers that propagate input optical signals from other switching nodes of photonic burst switching network 10 (Figure 10). The output leads of the optical demultiplexers

may be connected to the set of  $B$  core optical switches  $32_1$ - $32_B$  and to optical signal converter 36. For example, optical demultiplexer  $30_1$  has  $B$  output leads connected to input leads of the photonic burst switches  $32_1$ - $32_B$  (*i.e.*, one output lead of optical demultiplexer  $30_1$  to one input lead of each photonic burst switch) and at least one output lead connected to optical signal converter 36.

**[0041]** The output leads of photonic burst switches  $32_1$ - $32_B$  may be connected to optical multiplexers  $34_1$ - $34_A$ . For example, photonic burst switch  $32_1$  has  $A$  output leads connected to input leads of optical multiplexers  $34_1$ - $34_A$  (*i.e.*, one output lead of photonic burst switch  $32_1$  to one input lead of each optical multiplexer). Each optical multiplexer also has an input lead connected to an output lead of electrical-to-optical signal converter 38. Control unit 37 has an input lead or port connected to the output lead or port of optical-to-electrical signal converter 36. The output leads of control unit 37 may be connected to the control leads of photonic burst switches  $32_1$ - $32_B$  and electrical-to-optical signal converter 38. As described below in conjunction with the flow diagram of Figure 5, module 17 is used to receive and transmit optical control bursts, optical data bursts, and network management control labels. In one embodiment, the optical data burst and optical control bursts have transmission formats as shown in Figures 4A and 4B.

**[0042]** Figure 4A illustrates the format of an optical data burst for use in photonic burst switching network 10 (Figure 1), according to one embodiment of the present disclosure. In this embodiment, each optical data burst has a start guard band 40, an IP payload data segment 41, an IP header segment 42, a payload sync segment 43 (typically a small number of bits), and an end guard band 44 as shown in Figure 4A. IP payload data segment 41 includes the statistically-multiplexed IP data packets used to form the burst. Although Figure 4A shows the payload as contiguous, module 17 transmits payloads in a TDM format. Further, in some embodiments the data burst may be segmented over multiple variable-duration TDM channels. It should be pointed out that in this embodiment the optical data bursts and optical control bursts have local

significance only at the optical MAN, and may lose their significance at the optical WAN.

**[0043]** Figure 4B illustrates the format of an optical control burst for use in photonic burst switching network 10 (Figure 1), according to one embodiment of the present disclosure. In this embodiment, each optical control burst has a start guard band 46, an IP label data segment 47, a label sync segment 48 (typically a small number of bits), and an end guard band 49 as shown in Figure 4B. In this embodiment, label data segment 45 contains all the necessary routing and timing information of the IP packets or frames to form the optical burst. Although Figure 4B shows the control burst as contiguous, in this embodiment module 17 transmits control bursts in a TDM scheme.

**[0044]** In some embodiments, an optical network management control label (not shown) is also used in photonic burst switching network 10 (Figure 1). In such embodiments, each optical network management control label includes: a start guard band similar to start guard band 46; a network management data segment similar to data segment 47; a network management sync segment (typically a small number of bits) similar to label sync segment 48; and an end guard band similar to end guard band 44. In this embodiment, network management data segment contains network management information needed to coordinate transmissions over the network. In some embodiments, the optical network management control label is transmitted in a TDM scheme.

**[0045]** Figure 5 is a simplified block diagram architecture of certain portions of an edge router 500, that may be used as either an ingress node or egress node, according to one embodiment of the present disclosure. Various features are shown, such as a traffic shaper 501 and multiple data and control queues 510, 512, 514, 516. In other embodiment, such features may not be used or such functions may be implemented in other circuits. The edge router 500 may be utilized as either an ingress node 15 or an egress node 18 of the Figure 1 embodiment. In other embodiments, the portions shown in Figure 5 may

be used in a server, or in a server farm, that may again be used functionally as an ingress node or egress node. The edge router 500 may be used to support retransmission of control and data bursts when necessary.

5 [0046] IP packets (or frames) from multiple users and/or sources may arrive on signal line 540 for the traffic shaper module 501. The IP packets may be buffered in buffer 504 as necessary. The traffic shaper 501 may be used before the incoming IP packets are assembled to reduce or eliminate the traffic correlation structure and degree of self-similarity. 10 Thus, traffic burstiness reduction may improve contention resolution at the switching nodes and traffic performance. In one embodiment, the IP packets may be assembled into IP payload data 41 and label data 47 within data bursts and control bursts, respectively, by network processor ingress module 502. The network processor ingress module 502 may 15 then place the data bursts and control bursts in transmission order into the primary data queue 510 and primary control queue 514. The outputs of the queues may be selected by a switch 520 for transmission via framer 522. Leaving the framer 522 in electrical format, the data and control bursts may enter the optical MAN via the electrical to optical converter 20 524.

[0047] Data arriving from the optical MAN may enter edge router 500 via an optical to electrical converter 526. Once the data and control bursts are available to the framer 522 in electrical format, the framer 522 may supply them to network processor egress module 506 where they are 25 deframed and demultiplexed into IP packets.

[0048] In order to more conveniently provide for the retransmission of data and control bursts upon request, edge router 500 may also contain a secondary data queue 512 and secondary control queue 516. The secondary data queue 512 and secondary control queue 516 may contain 30 backup copies of data and control bursts. These backup copies may be needed when the copies of the data and control bursts within primary

data queue 510 and primary control queue 514 may be deleted upon successful transmission of the data bursts.

[0049] In one embodiment, edge router 500 may include a computer-readable media 570 which may include program code 572. This program  
5 code 572 may include updated operational software for use in network processor ingress module 502 and network processor egress module 506. Computer-readable media 570 may in some embodiments be a hard disk drive, a floppy disk drive, an optical disk drive, including compact disk read-only memory (CD-ROM) or versatile digital disk read-only memory  
10 (DVD-ROM), an electro-optical disk drive, a read-only memory (ROM) module, a flash memory module, or a network interface.

[0050] Figures 6A and 6B are diagrams illustrating a method of selecting a lightpath in a photonic burst switching network 600, according to one embodiment of the present disclosure. Figure 6A shows  
15 two ingress nodes 602, 606, two egress nodes 654, 656, and five switching nodes 610, 620, 630, 640, 650 for the sake of clarity. In other embodiments there may be many more ingress nodes, egress nodes, and switching nodes. In one embodiment, ingress nodes 602, 606 and egress nodes 654, 656 may be an edge router (i.e., LER) such as edge router 500  
20 of Figure 5. In other embodiments, ingress nodes 602, 606 and egress nodes 654, 656 may be implemented as a server, or as a server farm. Also shown are several optical fibers, 604, 608, 612, 614, 616, 622, 624, 632, 634, 644, 652, and 658 connecting the various nodes. Each of the optical fibers may support one or more lightpaths between ingress node  
25 602 and egress node 654. For example, photonic burst switching network 600, and specifically optical fiber 614, may support transmissions on three wavelengths,  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ . Let a lightpath segment be defined as any portion of a lightpath existing either between two switching nodes, between a switching node and an ingress node, or between a switching  
30 node and an egress node. Figure 6B shows schematically how three lightpath segments 660, 662, 664, corresponding to  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , respectively, may be carried by optical fiber 614. In other embodiments, there may be fewer than three or more than three wavelengths supported in a network.

[0051] In one embodiment, each lightpath segment may be assigned a weighting factor  $\alpha_{ij}(\lambda_k)$ , denoting that the weighting factor is assigned to the optical fiber from node  $i$  to node  $j$  on wavelength  $k$ . The use of a set of weighting factors may permit the determination of the likely lightpath  
5 segments, for transmitting a data burst that takes into consideration the network topology, network traffic statistical data, available network resources, as well as operator policies. These network traffic statistical data may include wavelength bandwidth utilization, class of service on each available wavelength, traffic load factors, and other data based upon  
10 observations of network traffic. In one embodiment, the set of weighting factors  $\alpha_{ij}(\lambda_k)$  may be predetermined and used as initialization values during the establishment of network operations.

[0052] In one embodiment, the weighting factors  $\alpha_{ij}(\lambda_k)$  may be modified as necessary during network operations to reflect changes in network  
15 traffic. For example, the weighting factor  $\alpha_{ij}(\lambda_k)$  of a heavily loaded lightpath segment may be reduced, in order to reduce future traffic loading, and to prevent traffic congestion or even loss of data on the heavily congested lightpath segments. The ingress node 602 may maintain a set of the weighting factors  $\alpha_{ij}(\lambda_k)$  to aid in its scheduling of the  
20 transmission of data bursts. In one embodiment the switching nodes 610, 620, 630, 640, 650 may track and maintain a record of network usage and network traffic statistics. From time to time the switching nodes 610, 620, 630, 640, 650 may send this network traffic statistical data to the ingress node 602 in a network management control label.  
25 Ingress node 602 may then use this data to determine updated weighting factors  $\alpha_{ij}(\lambda_k)$ . In other embodiments, the switching nodes 610, 620, 630, 640, 650 may determine the updated weighting factors  $\alpha_{ij}(\lambda_k)$  and transmit them to the ingress node 602.

[0053] The weighting factors  $\alpha_{ij}(\lambda_k)$  may be combined with other  
30 bandwidth reservation mechanisms to support the scheduling of data burst transmissions. In one embodiment, let  $B_{ij}(\lambda_k)$  denote the average

bit-rate between nodes  $i$  and  $j$ . Let  $B_i$  denote the average bit-rate among all the switching nodes between the  $i$ 'th pair of ingress and egress nodes, and among all wavelengths. Then a fine-grained bandwidth allocation factor  $f_{ij}(\lambda_k)$  may be calculated as:

$$f_{ij}(\lambda_k) = \alpha_{ij}(\lambda_k) [B_{ij}(\lambda_k) / B_i] \quad \text{Equation (1).}$$

The calculation may in one embodiment be performed in the ingress node 602. It may be useful to consider fine-grained bandwidth allocation factor  $f_{ij}(\lambda_k)$  as the fraction of the total bandwidth allocated between nodes  $i$  and  $j$  on wavelength  $k$  at any given time.

- 10 **[0054]** The fine-grained bandwidth allocation factors  $f_{ij}(\lambda_k)$  may be calculated for all of the lightpath segments on wavelength  $k$  at any given time in a given network. It may not be necessary to re-calculate the fine-grained bandwidth allocation factor  $f_{ij}(\lambda_k)$  until such time as one or more of the weighting factors  $\alpha_{ij}(\lambda_k)$  are updated. When an ingress node 602
- 15 considers the various lightpaths available for transmitting data bursts to the  $i$ 'th egress node 654, it may sum up the individual fine-grained bandwidth allocation factors  $f_{ij}(\lambda_k)$  of the component lightpath segments to form a coarse-grained bandwidth allocation factor  $f_i(\lambda_k)$  for the lightpath. The coarse-grained bandwidth allocation factor for wavelength  $k$   $f_i(\lambda_k)$  may
- 20 be calculated as:

$$f_i(\lambda_k) = [1/B_i] \sum_{j=1}^I \alpha_{ij}(\lambda_k) B_{ij}(\lambda_k) \quad \text{Equation (2).}$$

Here in Equation (2) the summation is over all the possible lightpath segments between ingress node 602 and egress node 654, which form a lightpath.

- 25 **[0055]** As an example, first assign index 1 to ingress node 602, index 2 to switching node 610, index 3 to switching node 630, index 4 to switching node 640, index 5 to switching node 650, and index 6 to egress node 654. Then consider a simple lightpath over optical fiber 616 between nodes with indices 2 and 5 and at wavelength  $\lambda_1$ . The coarse-

grained bandwidth allocation factor for this lightpath is simply  $f_2(\lambda_1) = [\alpha_{12}(\lambda_1)B_{12}(\lambda_1) + \alpha_{25}(\lambda_1)[B_{25}(\lambda_1) + \alpha_{56}(\lambda_1)[B_{56}(\lambda_1)]]/B_2$ . For a slightly more complicated lightpath 670 over optical fibers 604, 612, 634, 644, and 652 at wavelength  $\lambda_1$ , the coarse-grained bandwidth allocation factor for this

5 lightpath 670 is  $f_1(\lambda_1) = [\alpha_{12}(\lambda_1)[B_{12}(\lambda_1) + \alpha_{23}(\lambda_1)[B_{23}(\lambda_1) + \alpha_{34}(\lambda_1)[B_{34}(\lambda_1) + \alpha_{45}(\lambda_1)[B_{45}(\lambda_1) + \alpha_{56}(\lambda_1)[B_{56}(\lambda_1)]]]/B_3$ . For any combination of lightpath segments that form a lightpath between an ingress node and an egress node, the corresponding coarse-grained bandwidth allocation factor may similarly be calculated.

- 10 **[0056]** A scheduler within ingress switching node 602 may utilize the calculated values of the coarse-grained bandwidth allocation factors. In one embodiment the scheduler may be a portion of network processor ingress module 502. Due to transmission timing and synchronization uncertainties, the requested bandwidth may be generally larger than just
- 15 the data burst length. If the scheduler determines that there are one or more lightpaths with a high enough value of coarse-grained bandwidth allocation factor to meet the requested bandwidth, then the scheduler may select the lightpath which has the smallest coarse-grained bandwidth allocation factor with the same class of service as the requested
- 20 bandwidth.

- [0057]** In one embodiment, the scheduler may first assign a particular wavelength to the data burst. The assigned wavelength may be selected to match the transmission priority of the requested bandwidth for that data burst's length. The lightpaths using that wavelength may be
- 25 examined to see if one or more lightpaths possess a high enough value of coarse-grained bandwidth allocation factor to accept the given burst length. If there are one or more lightpaths, then the scheduler may schedule the transmission of the data burst along one of those lightpaths. If not, then an alternate wavelength may be considered. The lightpaths
- 30 using that alternate wavelength may then be examined to see if one or more lightpaths possess a high enough value of coarse-grained bandwidth allocation factor to meet the requested bandwidth. If there are one or more of these alternate lightpaths, then the scheduler may schedule the

transmission of the data burst along the lightpaths that first match the class of service of the requested bandwidth. If there is no lightpath that matches the priority of the requested bandwidth, then the scheduler may choose other available wavelengths.

- 5   **[0058]** If, however, there are no suitable lightpaths utilizing the alternate wavelength, then in a first embodiment the scheduler may select to delay the transmission of the requested bandwidth to a later time. The size of the delay time may be adjusted until a suitable lightpath is found.

- 10   **[0059]** In a second embodiment, the scheduler may select to break the data burst into two or more data burst segments. The scheduler may repeat the above process to determine whether there are lightpaths capable of accepting the data burst segments for transmission. In one embodiment the scheduler may seek the smallest number of such segments that may be transmitted. The transmitted segmented data  
15   burst maybe assembled into IP packets at the egress node 654.

- 20   **[0060]** In a third embodiment, the scheduler may consider lightpaths consisting of lightpath segments with different wavelengths. In this embodiment, some of the switching nodes may have the ability to receive data bursts on one wavelength, convert the data burst to another wavelength, and then send the data burst on the other wavelength. In some embodiments this may be accomplished using a tunable wavelength converter (TWC). Then lightpaths consisting of lightpath segments with different wavelengths may be examined to see if one or more lightpaths possess a high enough value of coarse-grained bandwidth allocation  
25   factor to accept the given burst length.

- 30   **[0061]** In one embodiment, the weighting factors  $\alpha_{ij}(\lambda_k)$  may be further dependent upon other factors. For example, the weighting factors  $\alpha_{ij}(\lambda_k)$  may be dependent upon offset time, where offset time is the time between a control burst and its corresponding data burst. In some embodiments a high priority data burst may be assigned a larger offset time. Therefore a weighting factor  $\alpha_{ij}(\lambda_k)$  corresponding to the larger offset may be increased.

**[0062]** In some embodiments, the method described above may be used in combination with other bandwidth reservation mechanisms that may be well-known in the art. Examples of such bandwidth reservation mechanisms may be a just-in-time JIT mechanism, a reserve-a-limited-duration RLD mechanism, or a reserve-a-fixed-duration RFD mechanism. A JIT mechanism may be used when the burst length is unknown, and only the starting time is known. A trailing control packet may be used to end the burst. In a RLD mechanism, only the expected end time of the burst is known, and bandwidth reservation may start immediately upon the arrival of the control packet. In the RFD mechanism, both the data burst start time and end time may be known, and the latest unused data burst channel may be selected.

**[0063]** Figure 7 is a simplified block diagram of software modules 700 for use in an ingress node or egress node such as the edge router shown in Figure 5, according to one embodiment of the present disclosure. In other embodiments the software modules 700 may be used in a server or in a server farm. The software modules 700 may process data packets in electronic form on input 750 connected to input buffer 702. The data packets may then be processed in the data packet parser 706. In one embodiment, the data packet parser 706 may deframe the incoming data packets, and then parse the header and other routing information from the data packets. The burstifier 734 may generate control bursts and data bursts from the received control and data packets from the parser 706. The control burst information may also include the data burst size, burst start and arrival times, arrival wavelength, source and destination addresses, forward equivalent class (FEC), burst priority, control burst offset time, and other relevant routing information. In another embodiment, the data packet parser 706 may classify the incoming data packets based on the extracted information before sending them to burst scheduler 730.

**[0064]** In another embodiment, the data packet parser 706 may additionally perform a cyclic redundancy check (CRC) on the incoming data packet. If the CRC fails on a particular data packet, then the data

packet parser 706 may generate a negative acknowledgement (NACK) for transmission back to the originator of the data packet. The originator upon receiving a NACK may then retransmit the failing control burst.

**[0065]** In one embodiment a synchronizer 704 may be connected  
5 between the output of the burst scheduler 730 and the queue manager 718. Synchronizer 704 may determine the synchronization of data packets by examining the time stamp within the data packet and comparing it to a local clock. Synchronizer 704 may then maintain synchronization between the incoming data packets and the newly-  
10 generated outgoing control bursts and data bursts.

**[0066]** Another function of the synchronizer 704 may in one embodiment be to maintain global synchronization with neighboring switching nodes as well as the ingress and/or egress nodes. In one embodiment such synchronization may be maintained by sending special  
15 sync messages to the neighboring switching nodes. The neighboring switching nodes may respond to the special sync messages by returning an acknowledgement message.

**[0067]** Finally, in one embodiment synchronizer 704 may assist in initializing optical MAN 11 of Figure 1. It may do this by initializing and  
20 maintaining synchronization of software modules 700 with the routers on the edges of optical MAN 11, such as ingress nodes 15 or egress nodes 18.

**[0068]** Burst scheduler 730 may schedule the transmission of data bursts to the next switching node or along a lightpath to an egress node. The scheduling may be accomplished using the method described above  
25 in connection with Figures 6A and 6B.

**[0069]** In one embodiment the network management controller 710 may be responsible for acting on the network configuration messages exchanged among the various ingress nodes, egress nodes, and switching nodes. As part of this functionality, network management controller 710  
30 may maintain a lookup table of current status, including list of the available variable duration TDM channels, source and destination addresses, wavelength usage, overall network traffic conditions, and

updated status of available network resources. Network management controller 710 may also control the filling of the input buffer 702 by sending a NACK back to the source of the data packets.

5 [0070] The forwarding engine 714 may gather the necessary control information produced in the network management controller 710 and burst scheduler 730. The forwarding engine 714 in one embodiment may select the next switching node through which to route the control and data bursts. This selection may be based in part upon destination address information within the control burst concerning the label-  
10 switched path (LSP). The forwarding engine 714 may also perform control label swapping. Control label swapping (routing information on a given wavelength for the next switching node) may be based upon updated destination addresses and the selected wavelength.

[0071] Using information presented by the forwarding engine 714, the  
15 control burst generator 716 in one embodiment produces the outgoing control bursts as established by the burst scheduler 730. The control burst generator 716 may also produce the control bursts for any PBS network configuration messages produced by the network management controller 710.

20 [0072] Finally, a queue manager 718 may produce a time sequenced output stream of control bursts. The outgoing control bursts may be sorted according to their relative priorities, destination addresses, and other network management priorities. During times of high traffic load, queue manager 718 may place lower-priority control bursts and network  
25 management labels into an output buffer 720 for temporary storage. In other embodiments, control bursts with a relatively longer time offset between control burst and data burst may also be stored within output buffer 720 when required.

[0073] Figure 8 is a flowchart diagram of a method discussed in  
30 connection with Figures 6A and 6B, according to one embodiment of the present disclosure. In one embodiment, the method of Figure 8 may be

performed in an ingress node, which may be the edge router 500 of Figure 5.

**[0074]** The process may begin with the initialization of a set of weighting factors  $\alpha_{ij}(\lambda_k)$  for all transmission wavelengths in block 804.

5 These weighting factors  $\alpha_{ij}(\lambda_k)$  may be default values stored locally or they may be derived values from a previous use of the network. Then in decision block 808, the process determines whether a data burst is available that creates the need for scheduling a transmission with an associated requested bandwidth. If so, then decision block 808 exits  
10 along the YES path, and in block 812 a first wavelength is selected on which to attempt to schedule the transmission of the requested data burst. This first wavelength upon which to attempt to schedule the transmission may be selected to match a transmission priority of the requested bandwidth. Here some wavelengths may be reserved for  
15 higher-priority transmissions while other wavelengths may be reserved for lower-priority transmissions. However, if the result of decision block 808 is negative, then the decision block 808 exits along the NO path and returns to the input of decision block 808.

**[0075]** After the selection of a wavelength for consideration, the process  
20 then determines in decision block 816 whether or not new network statistical data has arrived. If so, then decision block 816 exits along the YES path and in block 820 new values of the weighting factors  $\alpha_{ij}(\lambda_k)$  may be calculated. However, if no new network statistical data has arrived, then decision block 816 exits along the NO path. In both cases the  
25 process moves on to block 824.

**[0076]** In block 824, the weighting factors  $\alpha_{ij}(\lambda_k)$  may then be used to calculate a set of fine-grained bandwidth allocation factors  $f_{ij}(\lambda_k)$  for the wavelength currently under consideration. Once the set of fine-grained bandwidth allocation factors  $f_{ij}(\lambda_k)$  are calculated, then in block 828 they  
30 may be used to calculate a set of coarse-grained bandwidth allocation

factors  $f_i(\lambda_k)$  for the various possible lightpaths of the wavelength currently under consideration.

[0077] Then in block 832 the requested bandwidth is compared to each of the various coarse-grained bandwidth allocation factors  $f_i$  corresponding to the wavelength currently under consideration. In decision block 836, it is determined whether or not at least one of these coarse-grained bandwidth allocation factors  $f_i$  corresponding to the wavelength currently under consideration may support the requested bandwidth. If so, then decision block 836 exits along the YES path, and in block 844 the lightpath is selected that may have the smallest value of coarse-grained bandwidth allocation factors  $f_i$  that supports the requested bandwidth.

[0078] However, if none of the coarse-grained bandwidth allocation factors  $f_i$  corresponding to the wavelength currently under consideration supports the requested bandwidth, then decision block 836 exits along the NO path. In decision block 840, it is determined whether or not one or more wavelength exists that have not yet been considered that may support the requested bandwidth. If so, then decision block 840 selects one of those wavelengths as the next wavelength under consideration, and exits along the YES path. The YES path from decision block 840 returns the process to decision block 816 for further comparisons.

[0079] However, if no wavelength exists that has not yet been considered, then decision block 840 exits along the NO path. Then in block 852 the wavelength among all those considered that had the largest value of a coarse-grained bandwidth allocation factors  $f_i$  is selected. In decision block 856 it is determined whether or not to attempt to schedule the transmission by inserting a delay time. If so, then decision block 856 exits along the YES path and returns to decision block 808. However, if not, then decision block 856 exits along the NO path, and in block 860 the process segments the data burst.

**[0080]** Both block 844 and block 860 lead to block 848, where the data burst transmission is scheduled on the selected wavelength, using the lightpath determined, or, in some cases, the segmentation determined, in the above process.

- 5 **[0081]** In the foregoing specification, the invention has been described with reference to specific embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are,  
10 accordingly, to be regarded in an illustrative rather than a restrictive sense.

**CLAIMS**

What is claimed is:

1. A method, comprising:  
determining a plurality of weighting factors for a each of a plurality  
5 of lightpath segments;  
calculating a plurality of coarse-grained bandwidth allocation  
factors for each of a plurality of lightpaths formed from said lightpath  
segments between an ingress node and an egress node, utilizing said  
plurality of weighting factors;  
10 comparing a requested bandwidth with each of said coarse-grained  
bandwidth allocation factors for each of said lightpaths; and  
selecting one of said lightpaths that satisfies said requested  
bandwidth based on one of said coarse-grained bandwidth factors.
2. The method of claim 1, further comprising preselecting a first  
15 wavelength for transmission that matches a transmission priority of said  
requested bandwidth.
3. The method of claim 2, further comprising utilizing a second  
wavelength that matches said transmission priority of said requested  
bandwidth when none of said coarse-grained bandwidth allocation factors  
20 corresponding to said first wavelength indicates support of said requested  
bandwidth.
4. The method of claim 3, further comprising segmenting a data  
burst of said requested bandwidth when none of said coarse-grained  
bandwidth allocation factors corresponding to said second wavelength  
25 that matches said transmission priority of said requested bandwidth.

5. The method of claim 3, further comprising delaying transmission of said data burst when none of said coarse-grained bandwidth allocation factors corresponding to said second wavelength indicates support of said requested bandwidth.

5 6. The method of claim 1, wherein said determining a plurality of weighting factors includes utilizing network traffic statistical data.

7. The method of claim 6, further comprising transferring said network traffic statistical data from a switching node.

10 8. The method of claim 1, wherein said calculating a plurality of coarse-grained bandwidth allocation factors includes summing a plurality of fine-grained bandwidth allocation factors.

9. The method of claim 8, further comprising forming said fine-grained bandwidth allocation factors by multiplying one of said weighting factors by a ratio of a mean bit rate corresponding to said one of said  
15 weighting factors to an overall mean bit rate.

10. An apparatus, comprising:  
means for determining a plurality of weighting factors for a each of a plurality of lightpath segments;  
means for calculating a plurality of coarse-grained bandwidth  
20 allocation factors for each of a plurality of lightpaths formed from said lightpath segments between an ingress node and an egress node, utilizing said plurality of weighting factors;  
means for comparing a requested bandwidth with each of said coarse-grained bandwidth allocation factors for each of said lightpaths;  
25 and

means for selecting one of said lightpaths that satisfies said requested bandwidth based on one of said coarse-grained bandwidth factors.

11. The apparatus of claim 10, further comprising means for  
5 preselecting a first wavelength for transmission that matches a transmission priority of said requested bandwidth.

12. The apparatus of claim 11, further comprising means for  
utilizing a second wavelength that matches said transmission priority of  
said requested bandwidth when none of said coarse-grained bandwidth  
10 allocation factors corresponding to said first wavelength indicates support of said requested bandwidth.

13. The apparatus of claim 12, further comprising means for  
segmenting a data burst of said requested bandwidth when none of said  
coarse-grained bandwidth allocation factors corresponding to said second  
15 wavelength indicates support of said requested bandwidth.

14. The apparatus of claim 13, further comprising means for  
delaying transmission of said data burst when none of said coarse-  
grained bandwidth allocation factors corresponding to said second  
wavelength indicates support of said requested bandwidth.

20 15. The apparatus of claim 10, wherein said means for  
determining a plurality of weighting factors includes means for utilizing  
network traffic statistical data.

16. The apparatus of claim 15, further comprising transferring  
said network traffic statistical data from a switching node.

25 17. The apparatus of claim 10, wherein said means for  
calculating a plurality of coarse-grained bandwidth allocation factors

includes means for summing a plurality of fine-grained bandwidth allocation factors.

18. The apparatus of claim 17, further comprising means for forming said fine-grained bandwidth allocation factors including means  
5 for multiplying one of said weighting factors by a ratio of a mean bit rate corresponding to said one of said weighting factors to an overall mean bit rate.

19. A computer-readable media, including programming code that when executed by a processor performs the following:  
10 determining a plurality of weighting factors for each of a plurality of lightpath segments;  
calculating a plurality of coarse-grained bandwidth allocation factors for each of a plurality of lightpaths formed from said lightpath segments between an ingress node and an egress node, utilizing said  
15 plurality of weighting factors;  
comparing a requested bandwidth with each of said coarse-grained bandwidth allocation factors for each of said lightpaths; and  
selecting one of said lightpaths that satisfies said requested bandwidth based on one of said coarse-grained bandwidth factors.

20. The computer-readable media of claim 19, where said programming code further comprises preselecting a first wavelength for transmission that matches a transmission priority of said requested bandwidth.

21. The computer-readable media of claim 20, where said  
25 programming code further comprises utilizing a second wavelength that matches said transmission priority of said requested bandwidth when none of said coarse-grained bandwidth allocation factors corresponding to said first wavelength indicates support of said requested bandwidth.

22. The computer-readable media of claim 21, where said programming code further comprises segmenting a data burst of said requested bandwidth when none of said coarse-grained bandwidth allocation factors corresponding to said second wavelength that matches  
5 said transmission priority of said requested bandwidth.

23. The computer-readable media of claim 21, where said programming code further comprises delaying transmission of said data burst when none of said coarse-grained bandwidth allocation factors corresponding to said second wavelength indicates support of said length  
10 of said requested bandwidth.

24. The computer-readable media of claim 19, wherein said determining a plurality of weighting factors includes utilizing network traffic statistical data.

25. The computer-readable media of claim 24, where said  
15 programming code further comprising transferring said network traffic statistical data from a switching node.

26. The computer-readable media of claim 19, wherein said calculating a plurality of coarse-grained bandwidth allocation factors includes summing a plurality of fine-grained bandwidth allocation factors.

20 27. The computer-readable media of claim 26, where said programming code further comprises forming said fine-grained bandwidth allocation factors by multiplying one of said weighting factors by a ratio of a mean bit rate corresponding to said one of said weighting factors to an overall mean bit rate.

28. An apparatus, comprising:  
a network processor ingress module to maintain a set of weighting  
factors and coarse-grained bandwidth allocation factors; and  
a network processor egress module to receive network statistical  
5 data to support said set of weighting factors.

29. The apparatus of claim 28, wherein said network processor  
ingress module determines said coarse-grained bandwidth allocation  
factors from said set of weighting factors.

30. The apparatus of claim 28, wherein said network processor  
10 ingress module updates said set of weighting factors based upon said  
network statistical data.

31. The apparatus of claim 28, wherein said network processor  
ingress module determines a lightpath for a requested bandwidth by  
comparing said requested bandwidth to said coarse-grained bandwidth  
15 allocation factors.

1/8

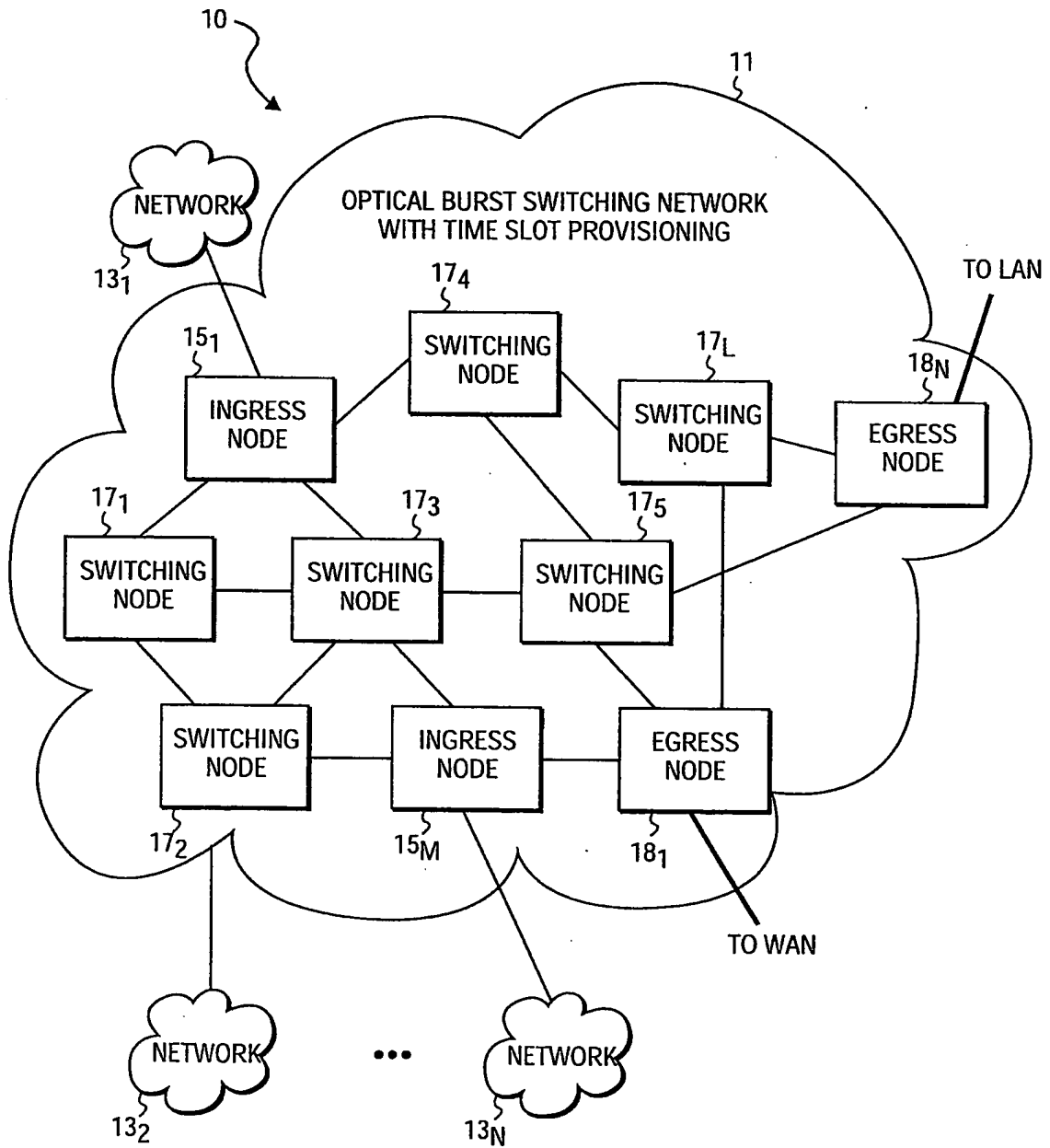


FIG. 1

2/8

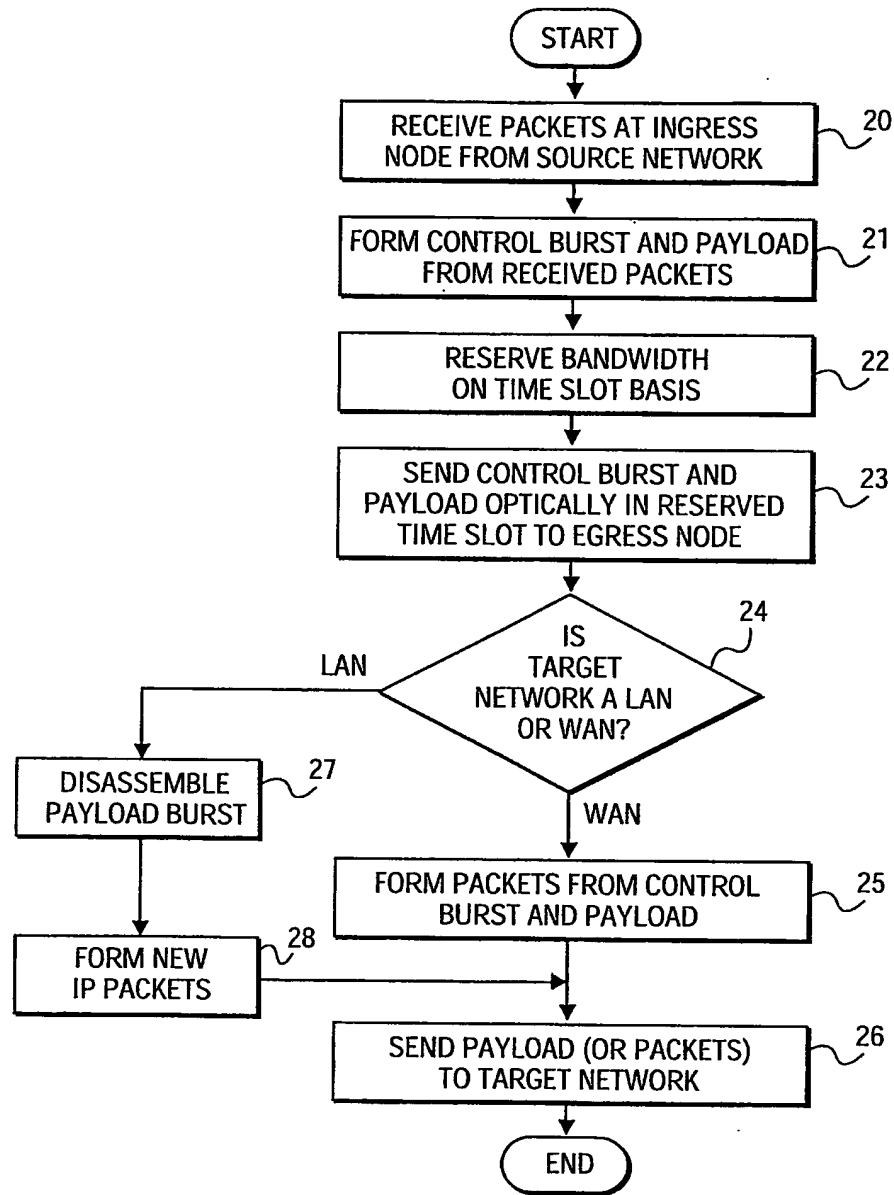


FIG. 2

3/8

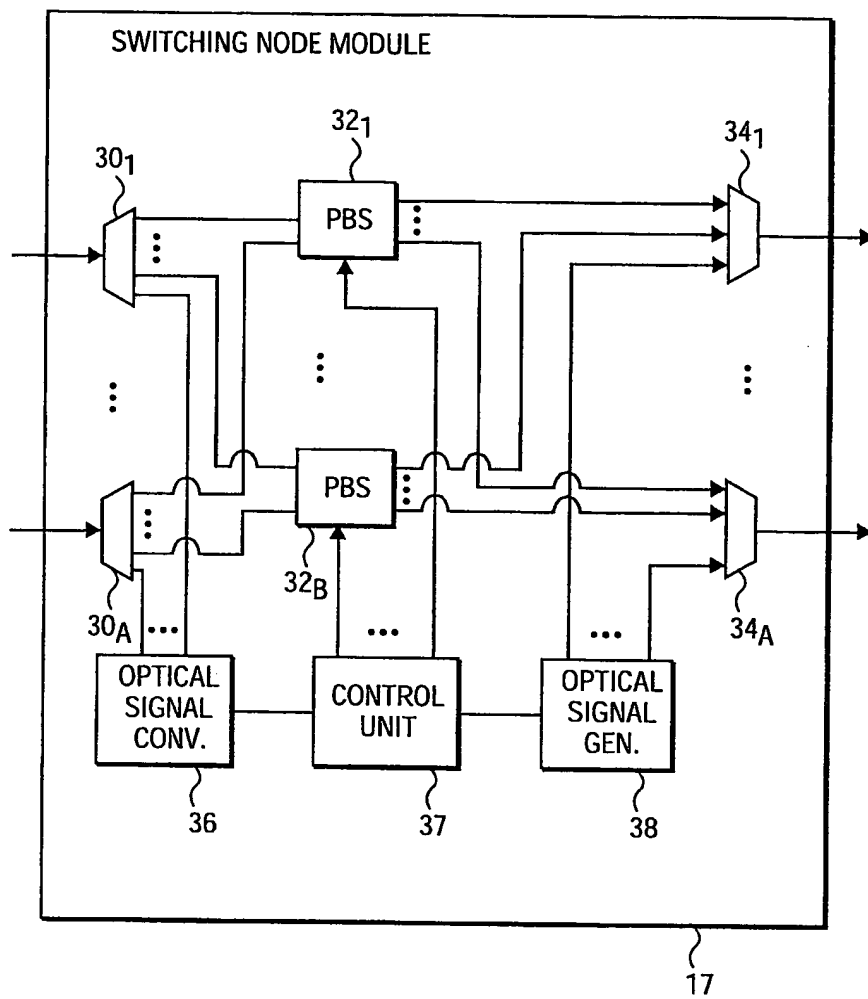


FIG. 3

4/8

## OPTICAL DATA BURST FORMAT

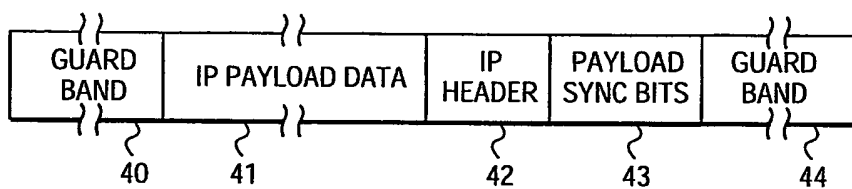


FIG. 4A

## OPTICAL CONTROL BURST FORMAT

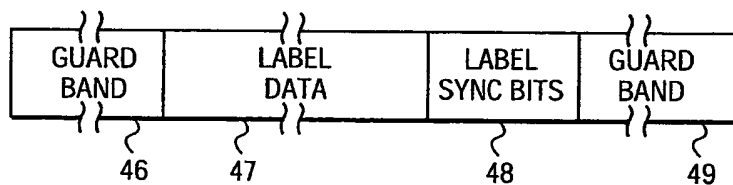


FIG. 4B

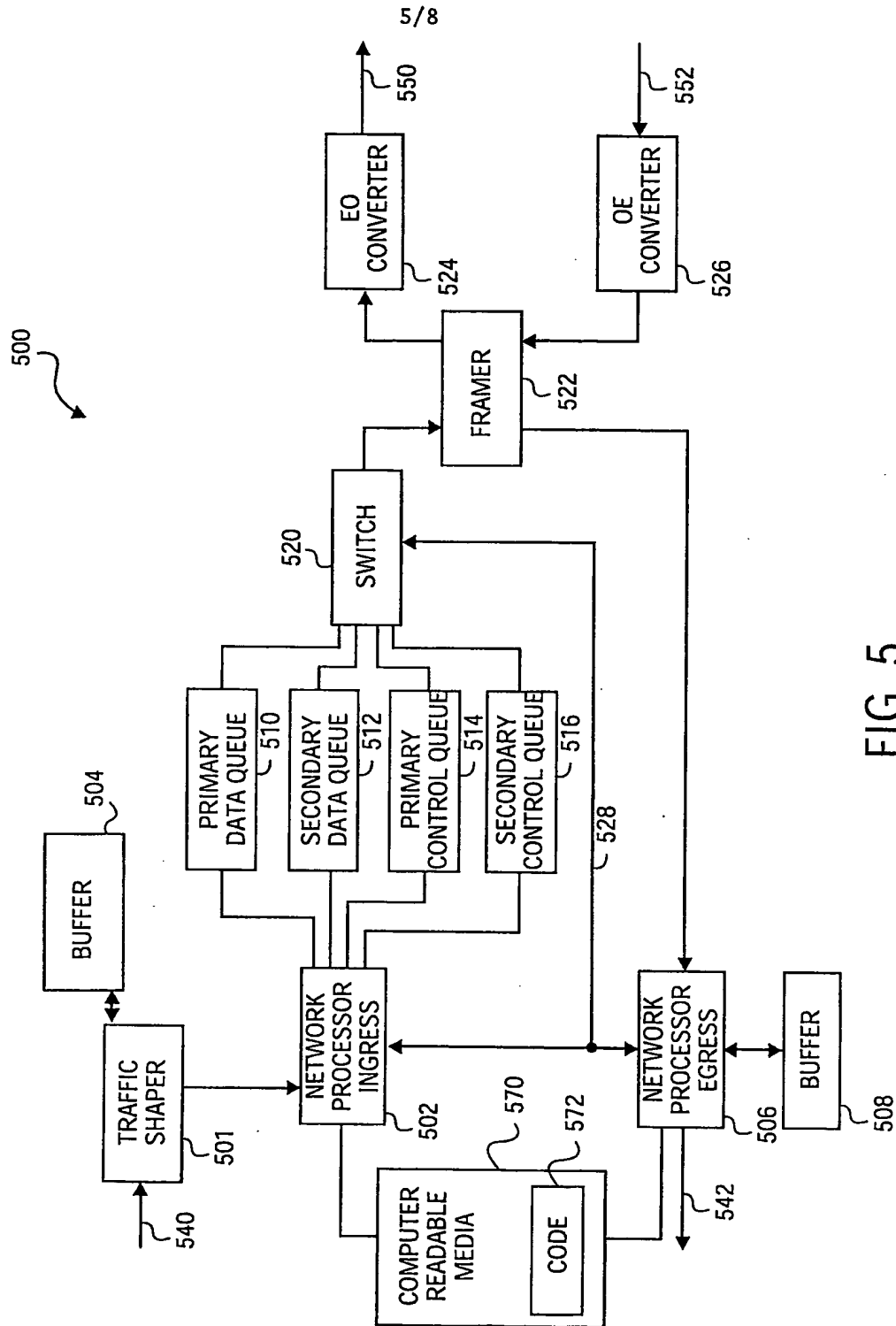


FIG. 5

6/8

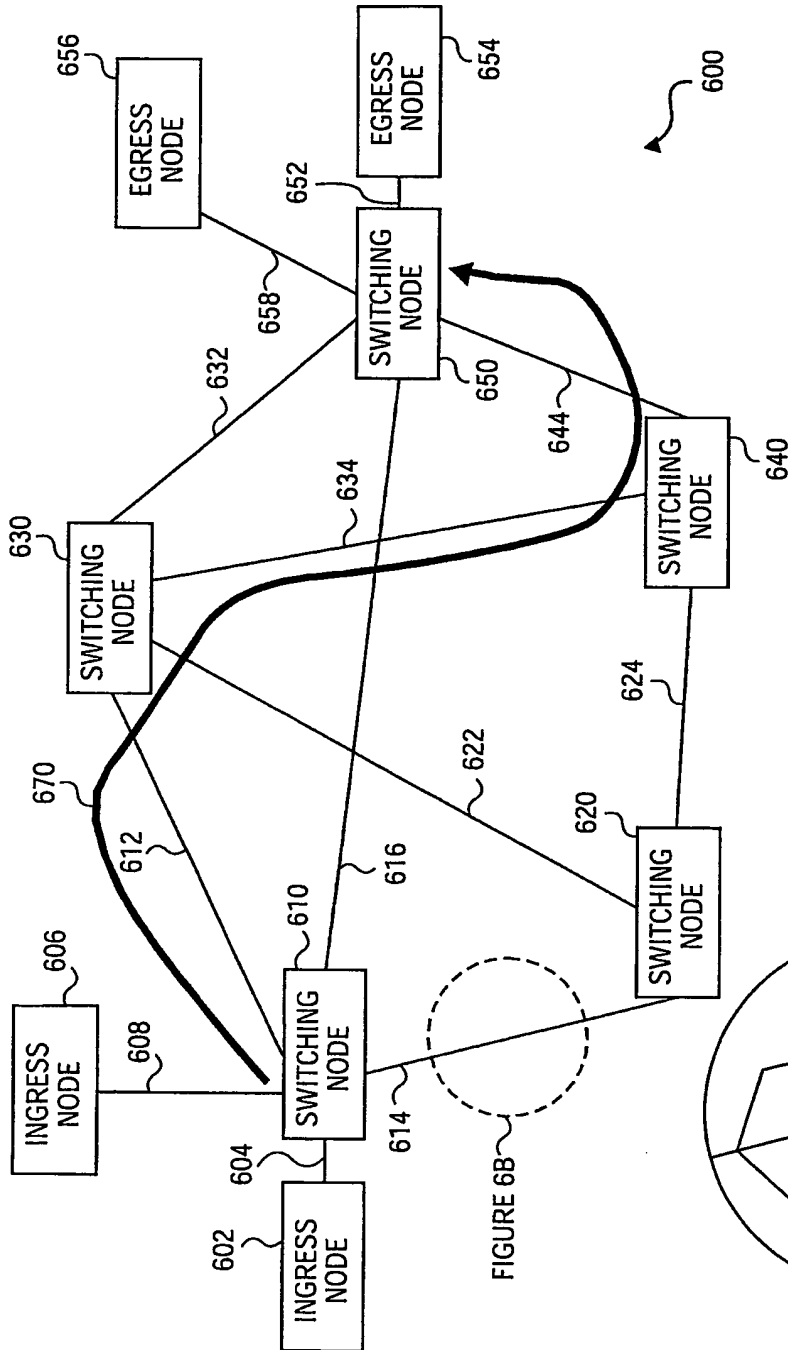


FIG. 6A

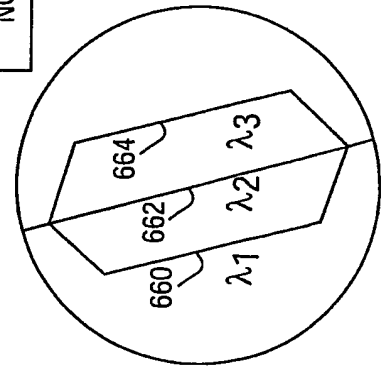


FIG. 6B

7/8

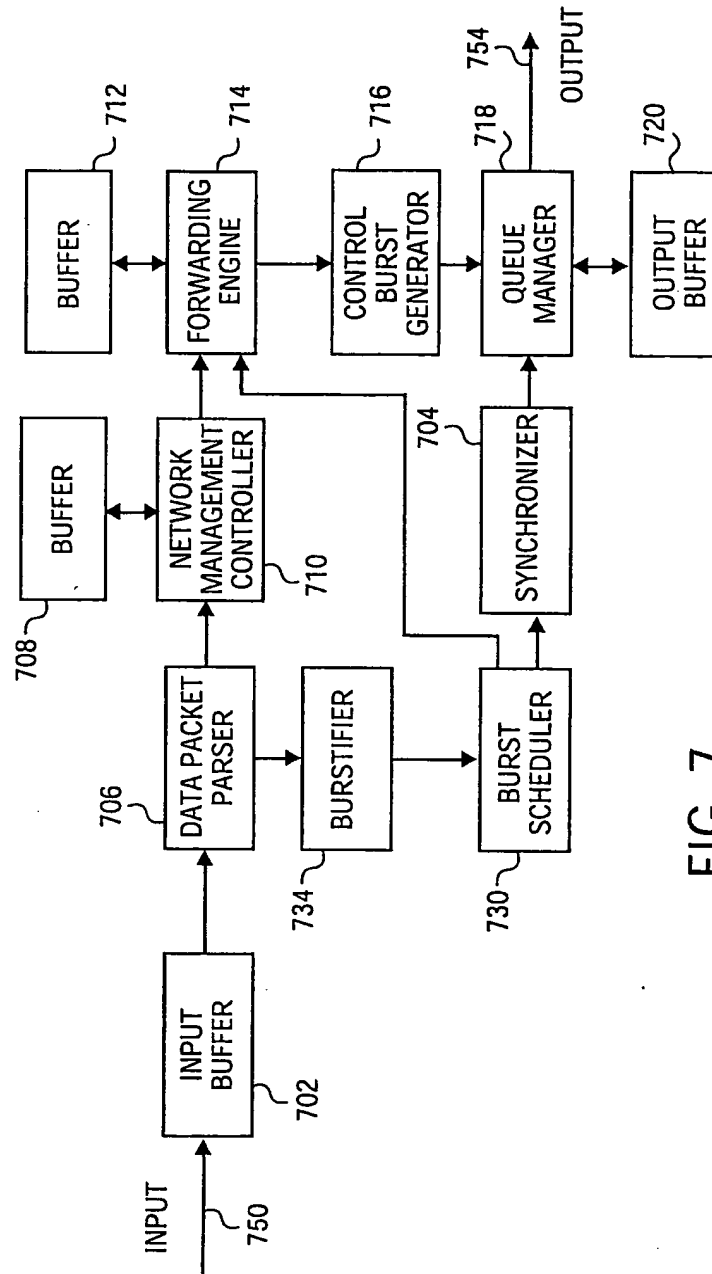
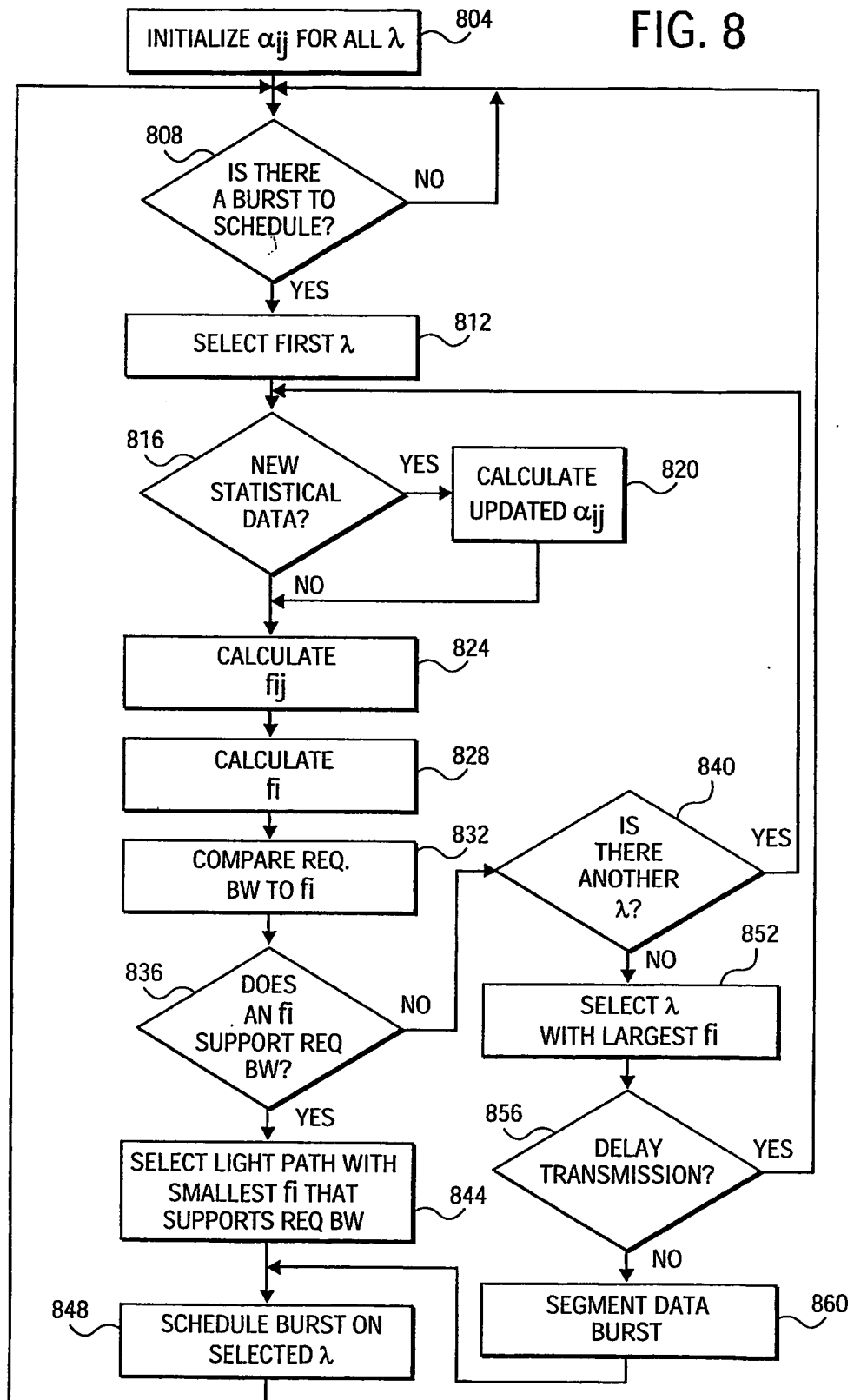


FIG. 7

8/8

FIG. 8



# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US 03/38613

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 H04Q11/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 H04L H04Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages                                   | Relevant to claim No. |
|------------|--|-----------------------|
| P,X        | EP 1 303 111 A (CIT ALCATEL)<br>16 April 2003 (2003-04-16)<br>column 4, paragraph 0012 - column 5,<br>paragraph 0019 | 1,10,19,<br>28        |
| P,X        | US 2003/009582 A1 (QIAO CHUNMING ET AL)<br>9 January 2003 (2003-01-09)<br>paragraph '0090! - paragraph '0096!        | 1,10,19,<br>28        |
| A          | EP 1 135 000 A (ERICSSON TELEFON AB L M)<br>19 September 2001 (2001-09-19)<br>the whole document                     | 1-31                  |
| A          | EP 0 876 076 A (NIPPON ELECTRIC CO)<br>4 November 1998 (1998-11-04)<br>column 4, line 49 - column 5, line 56         | 1-31                  |

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

28 April 2004

Date of mailing of the international search report

07/05/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Meurisse, W

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 03/38613

| Patent document<br>cited in search report |    | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|---|----|---------------------|----------------------------|---------------------|
| EP 1303111                                | A  | 16-04-2003          | CN 1412968 A               | 23-04-2003          |
|   |    |                     | EP 1303110 A2              | 16-04-2003          |
|   |    |                     | EP 1303111 A2              | 16-04-2003          |
|   |    |                     | US 2003067880 A1           | 10-04-2003          |
| US 2003009582                             | A1 | 09-01-2003          | WO 03003156 A2             | 09-01-2003          |
| EP 1135000                                | A  | 19-09-2001          | EP 1135000 A1              | 19-09-2001          |
|   |    |                     | US 2001030962 A1           | 18-10-2001          |
| EP 0876076                                | A  | 04-11-1998          | JP 3063721 B2              | 12-07-2000          |
|   |    |                     | JP 11015804 A              | 22-01-1999          |
|   |    |                     | CA 2236384 A1              | 30-10-1998          |
|   |    |                     | EP 0876076 A2              | 04-11-1998          |
|   |    |                     | US 6385201 B1              | 07-05-2002          |